PREFACE

If you have picked up this book you surely love sports and you probably like math. You may have read Michael Lewis's great book *Moneyball*, which describes how the Oakland A's used mathematical analysis to help them compete successfully with the New York Yankees even though the average annual payroll for the A's is less than 40 percent of that of the Yankees. After reading *Moneyball*, you might have been curious about how the math models described in the book actually work. You may have heard how a former night watchman, Bill James, revolutionized the way baseball professionals evaluate players. You probably want to know exactly how James and other "sabermetricians" used mathematics to change the way hitters, pitchers, and fielders are evaluated. You might have heard about the analysis of Berkeley economic professor David Romer that showed that NFL teams should rarely punt on fourth down. How did Romer use mathematics to come up with his controversial conclusion? You might have heard how Mark Cuban used math models (and his incredible business savvy) to revitalize the moribund Dallas Mavericks franchise. What mathematical models does Cuban use to evaluate NBA players and lineups? Maybe you bet once in a while on NFL games and wonder whether math can help you do better financially. How can math determine the true probability of a team winning a game, winning the NCAA tournament, or just covering the point spread? Maybe you think the NBA could have used math to spot Tim Donaghy's game fixing before being informed about it by the FBI. This book will show you how a statistical analysis would have "red flagged" Donaghy as a potential fixer.

If *Moneyball* or day-to-day sports viewing has piqued your interest in how mathematics is used (or can be used) to make decisions in sports and sports gambling, this book is for you. I hope when you finish reading the book you will love math almost as much as you love sports.

To date there has been no book that explains how the people running Major League Baseball, basketball, and football teams and Las Vegas sports bookies use math. The goal of *Mathletics* is to demonstrate how simple arithmetic, probability theory, and statistics can be combined with a large dose of common sense to better evaluate players and game strategy in America's major sports. I will also show how math can be used to rank sports teams and evaluate sports bets.

Throughout the book you will see references to Excel files (e.g., Standings.xls). These files may be downloaded from the book's Web site, http://www.waynewinston.edu).

Mathletics by Winston

use this chapter w/ regression

41

WHICH LEAGUE HAS GREATER PARITY, THE NFL OR THE NBA?

Sports fans love the NFL because it seems like there is always a surprise team that wins the Super Bowl or challenges for the championship. For example, who expected Tampa Bay to win the Super Bowl in 2002? NBA fans complain the same teams (such as Detroit and San Antonio) are always on top. It is easy to show that the NFL does indeed exhibit more parity and unexpected team performances than does the NBA.

If a league has a great deal of parity you would expect it to be difficult to predict a team's performance based on their previous year's performance. That is, teams that do poorly one season should have a good shot at being above average the following season, and vice versa. If a league exhibits little parity you would expect that it would be relatively easy to predict a team's performance for one season using the previous year's performance. What metric should we use to measure team performance? The simplest metric would be regular season wins, but we have seen that NFL teams play schedules that differ by 7 or more points in difficulty. A team with a tough schedule and 9–7 record might be much better than a team that went 11-5 with an easy schedule. We will use the final Sagarin rating as our metric for team performance during a season.[1] These ratings include all regular and post-season games and are quite similar to the least squares ratings described in chapter 40. The file Parity.xls contains the ratings for each NFL team and NBA team for five seasons (for the NBA, 2002–3 through 2006–7 seasons and for the NFL, the 2002–6 seasons). We try to predict each team's Sagarin rating during a year based on their previous year's Sagarin rating. For example, for our NBA data we would have the following four data points for the Spurs: (Spurs 2002–3 rating,

[1] See http://www.usatoday.com/sports/sagarin.htm.

Spurs 2003–4 rating), (Spurs 2003–4 rating, Spurs 2004–5 rating), (Spurs 2004–5 rating, Spurs 2005–6 rating), and (Spurs 2005–6 rating, Spurs 2006–7 rating). We simply find the best-fitting line to predict the following season's rating based on the previous season's rating. Sagarin ratings average to 20 in the NFL and 90 in the NBA.

Figures 41.1 and 41.2 give the results obtained using the Excel Trend Curve feature. We find that the following year's NFL team performance has a .35 correlation with the team's performance the previous year, while the following year's NBA team performance has a .56 correlation with the team's performance the previous year. This shows that it is much easier to predict an NBA team's performance than an NFL team's performance us-

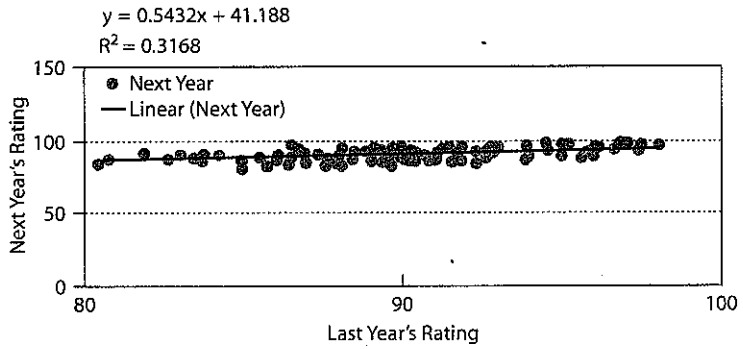

Figure 41.1. Predicting NBA team performance. The correlation between the previous year's Sagarin rating and the following year's rating is obtained by taking the square root of the R Squared values.

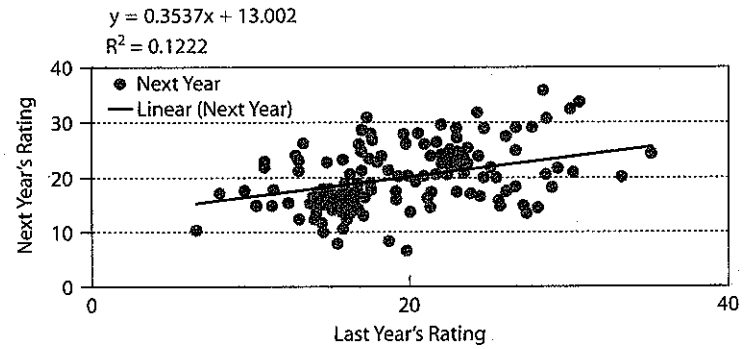

Figure 41.2. Predicting NFL team performance.

ing the previous year's data. This means that good NBA teams are much more likely to stay good than are good NFL teams. Conversely, bad NFL teams have a much better shot at being good the following year than do bad NBA teams. This insight is consistent with the prevalent view that the NFL has more parity than the NBA. Looking at the least squares lines shown in figures 41.1 and 41.2, we would predict that an NBA team that is 10 points better than average during a season (100 rating) would be predicted to have a rating of $41.19 + .5432(100) = 95.51$ the following year (5.51 points better than average), while an NFL team that performed 10 points better than average (30 rating) during a season would be predicted to have a rating of $13 + .3537(30) = 23.61$ (3.61 points better than average). Similarly, an NBA team that is 10 points worse than average would be predicted to play 5.51 points worse than average the following year and an NFL team that played 10 points worse than average would be expected to play only 3.61 points worse than average the following year. Thus for each league we predict a team to play closer to average the following year than they did the previous year. This is an example of regression toward the mean. NFL teams regress more to the mean than do NBA teams. The fact that a good NFL team on average ends up closer to average the following season than does a good NBA team must mean that more good NFL teams play poorly the following season (and more bad NFL teams play well).

Possible Explanations for NFL Parity

Why does the NFL exhibit more parity than the NBA? The NFL has a "hard" salary cap ($109 million per team in 2007), which teams cannot exceed. The NBA penalizes teams that exceed the salary cap ($55.6 million in 2007), but teams are allowed to exceed the salary cap. If an NFL team performs well then their star players may seek higher salaries elsewhere because the hard cap prevents a great team from rewarding all their good players.

In addition, in the NFL draft the worst team is guaranteed the first draft pick. In the NBA a lottery is held and the worst team has only a 25% chance of getting the first pick. In short, the NFL draft has teams draft in inverse order of team performance while the NBA only approximates teams drafting in inverse order of performance. Therefore, it seems reasonable to conclude that the NFL draft would create more parity than would the NBA draft.

The NBA and the NFL also differ in terms of the way player contracts are handled. In the NFL no contract is guaranteed. A player can have a six-year, $100 million contract and if the team cuts the player then the team does not have to pay the player. In the NBA, however, most player contracts are structured so that if a team cuts a player the team still owes the player his money. This dichotomy makes it much easier to change an NFL roster than change an NBA roster. Since it is harder to change an NBA team roster, we would expect that NBA teams' performance during the following season would be more similar to their performance the previous season than would be the case with the performance of an NFL team.

These three differences between the NBA and the NFL appear to be the major contributors to the greater degree of parity exhibited by the NFL.

42

THE RATINGS PERCENTAGE INDEX (RPI)

During the college basketball season hoop fans anxiously anticipate selection and seeding of teams for the NCAA tournament. The NCAA lection committee wants an accurate view of the teams' relative abili but, like the BCS selection committee, the NCAA tournament selec committee wants to use only a team's win-loss record and not the scor their games to rank teams. The NCAA believes that including game sc in the ranking and seeding process would cause the top teams to try run up the score on lesser opponents. In chapter 40 we explained hov use a logistic regression-based ranking system to rank NFL teams. An ic tical system would do an excellent job of ranking college basketball tea Defying logic, however, the NCAA uses the complex and flawed Rating centage Index (RPI) to rank college basketball teams.

Let's suppose we want to compute the RPI of Indiana University (I IU's RPI ranking is computed as a weighted average of three quantitie

- IU's own winning percentage (referred to as TWP).
- Not counting the games involving IU, the winning percentage of each o IU's opponents. These winning percentages are averaged to compute OPP = opponent's average winning percentage.
- OPPOPP = the average winning probability of IU's opponents' opponent (not including games in which they play any of IU's opponents but includ ing games played against IU).

The home team in college basketball wins around 70% of the games tween equally matched teams. This gave teams like Duke, which pla many more home games than road games, an unfair advantage in the rankings. Beginning with the 2005 season, a home win or road loss counted as 0.6 wins or 0.6 losses. An away win or home loss was cou